Big Data and Computing Visions

www.bidacv.com

Big. Data. Comp. Vis. Vol. 1, No. 3 (2021) 140–145.

Paper Type: Original Article

Credit Card Fraud Detection Through Machine Learning Algorithm

Agyan Panda[1,*], Bharath Yadlapalli[2], Zhi Zhou[3]

[1] Department of Computer Science and Engineering, OEC Engineering College, OD, India; hmpendrive@gmail.com.
[2] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, AP, India.
[3] Government Information Headquarters Inspur Software Group Company Ltd, Jinan, China.

Citation:

Panda, A., Yadlapalli, B., & Zhou, Z. (2021). Credit card fraud detection through machine learning algorithm. *Big data and computing visions, 1* (3), 140-145.

Received: 15/04/2021 Reviewed: 22/05/2021 Revised: 13/06/2021 Accept: 12/07/2021

Abstract

Every year, millions of dollars are lost due to fraudulent credit card transactions. To help fraud investigators, more algorithms are turning to powerful machine learning methodologies. Designing fraud detection algorithms is particularly difficult because to the non-stationary distribution of data, excessively skewed class distributions, and continuous streams of transactions. At the same time, due to confidentiality considerations, public data is uncommon, leaving many questions unanswered about the best technique for dealing with them. We present some replies from the practitioners in this publication. Un balanced ness, non- stationarity and assessment. Our industrial partner provided us with an actual credit card dataset, which we used to do the analysis. In this project, we attempt to develop and evaluate a model for the imbalanced credit card fraud dataset.

Keywords: Credit card fraud, Machine learning applications, Data science, Automated fraud detection.

1 | Introduction

Licensee **Big Data and Computing Vision**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

In credit card transactions, fraud is defined as the unlawful and unwelcome use of an account by someone other than the account owner. To stop this misuse, necessary preventative steps should be adopted, and the behaviour of such fraudulent operations can be studied to reduce it and guard against its similar occurrences in the future. People have been concerned with credit card fraud detection models based on data mining in recent years. Classic data mining algorithms aren't directly applicable to our topic because it's handled as a classification challenge. As a result, a different technique is employed, which involves the employment of general-purpose meta heuristics like genetic algorithms. The enhancement of science and technology leads to make the life more comfortable than older days. The emerging technologies like software engineering [1] and [2], energy management [3], [4] and [5], wireless sensor network [6]-[13], face recognition [14], neural network [14], routing [15] and [16], cloud computing [17], distributive environment [18], mixed environment [19], bellman algorithm [20], programming language [21], neutrosophic shortest path [22], [23] and [24], optimal path [25], multi-objective optimal path [26], transportation problem [27], [28] and [29], uncertainty problem [30]-[35]

Corresponding Author: hmpendrive@gmail.com

10.22105/BDCV.2021.142231

, fuzzy shortest path [36] and [37], answer note [38], making the products more intelligent and self-healing based. The smart city [39] and [40] applications like smart water [40] and [41], smart agriculture [42] smart grid [42] and [43], smart parking [44], smart resource management, etc. are based on IoT [44] and [45] and IoE technologies. The purpose of this study is to create a credit card fraud detection system based on genetic algorithms. Genetic algorithms are a form of evolutionary algorithm that tries to continuously improve solutions. When a card is duplicated, stolen, or lost by fraudsters, it is usually utilized until the available limit is exhausted. As a result, rather than focusing on the quantity of correctly classified transactions, a strategy that reduces the overall allowed limit on fraud-prone cards takes precedence. Its goal is to reduce false alerts by utilizing a genetic algorithm to optimize a set of interval-valued parameters.

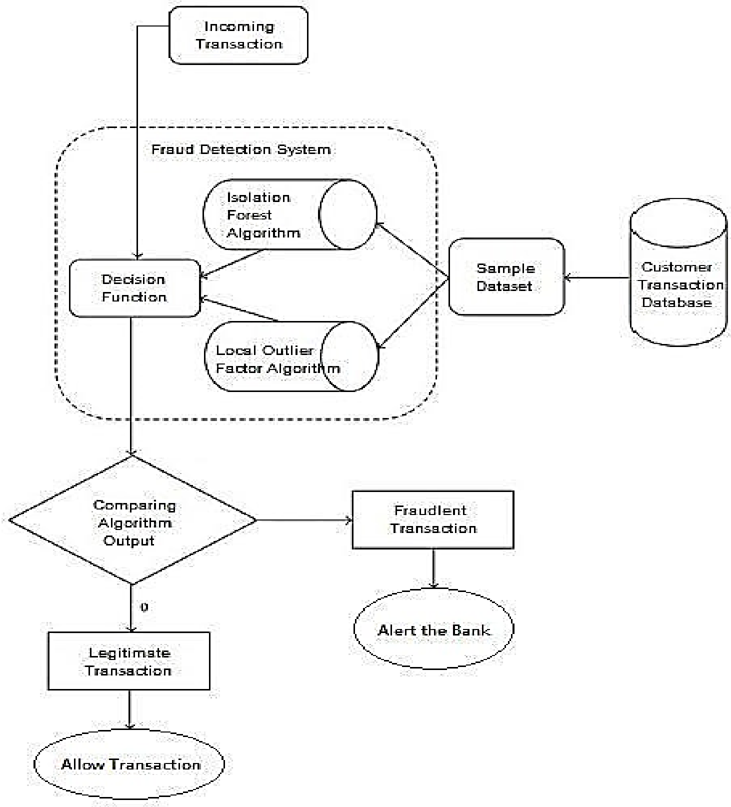

Fig. 1. Flowchart of the proposed model.

In recent years, people have been concerned about credit card fraud detection methods based on data mining. Classic data mining algorithms aren't directly applicable to our topic because it's handled as a classification challenge. As a result, a different technique is employed, which involves the employment of general-purpose meta heuristics like genetic algorithms. The purpose of this study is to create a credit card fraud detection system based on genetic algorithms. Genetic algorithms are a form of evolutionary algorithm that tries to continuously improve solutions. When a card is duplicated, stolen, or lost by fraudsters, it is usually utilized until the available limit is exhausted. As a result, rather than focusing on the quantity of correctly classified transactions, a strategy that reduces the overall allowed limit on fraud-prone cards takes precedence. Its goal is to reduce false alerts by utilizing a genetic algorithm to optimize a set of interval-valued parameter.

2 | Literature Review

Traditionally, fraud detection has been thought of as a data mining task aimed at correctly identifying transactions as lawful or fraudulent. Many performance indicators are defined for classification problems, the most of which are connected to the correct number of cases categorized correctly. A more appropriate solution is necessary due to the inherent structure of credit card transactions. When a card is duplicated, stolen, or lost by fraudsters, it is usually utilized until the available limit is exhausted. As a result, rather

than focusing on the quantity of correctly classified transactions, a strategy that reduces the overall allowed limit on fraud-prone cards takes precedence.

Panda and Yadlapalli| Big. Data. Comp. Vis. 1(3) (2021) 140-145

3 | Methodology

The method proposed in this research employs the most up-to-date machine learning methods to detect aberrant activity known as outliers. The following is a representation of the fundamental rough architecture diagram. First and foremost, we got our data from Kaggle, a data analysis service that offers datasets. There are 31 columns in this dataset, with 28 of them labelled v1-v28 to preserve sensitive information. The other columns represent Time, Amount and Class. The time difference between the first and subsequent transactions is shown in this graph. Amount is the amount of money transacted. Class 0 represents a valid transaction and 1 represents a fraudulent one We use a variety of graphs to visually inspect the dataset for discrepancies.

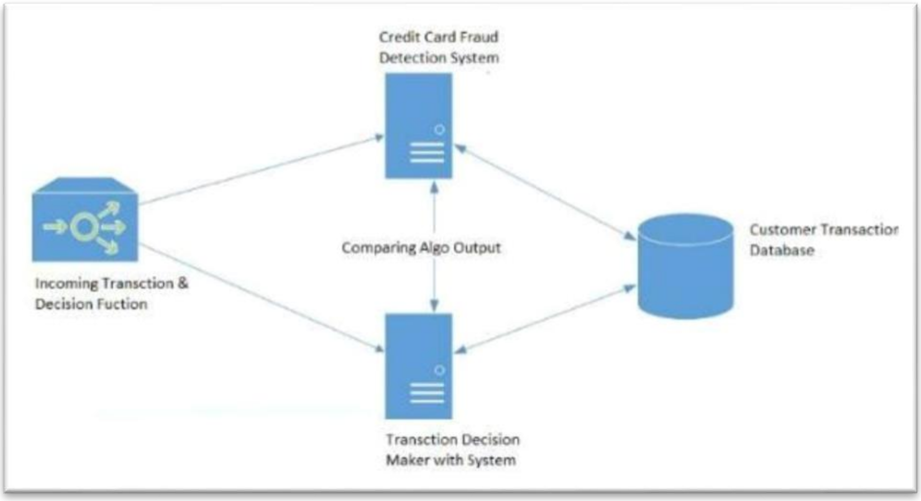

Fig. 2. Functional diagram of the proposed model.

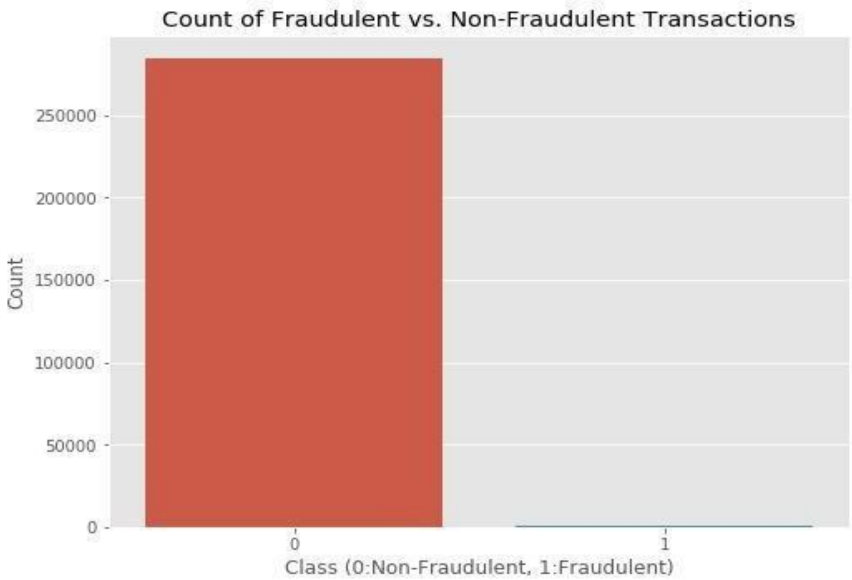

Fig. 3. Performance statistics.

The number of fraudulent transactions is substantially fewer than the number of valid transactions, as seen in this graph. Following this, we create a heatmap to visualise the data in colour and investigate the relationship between our predictive factors and the class variable. This heatmap is shown below

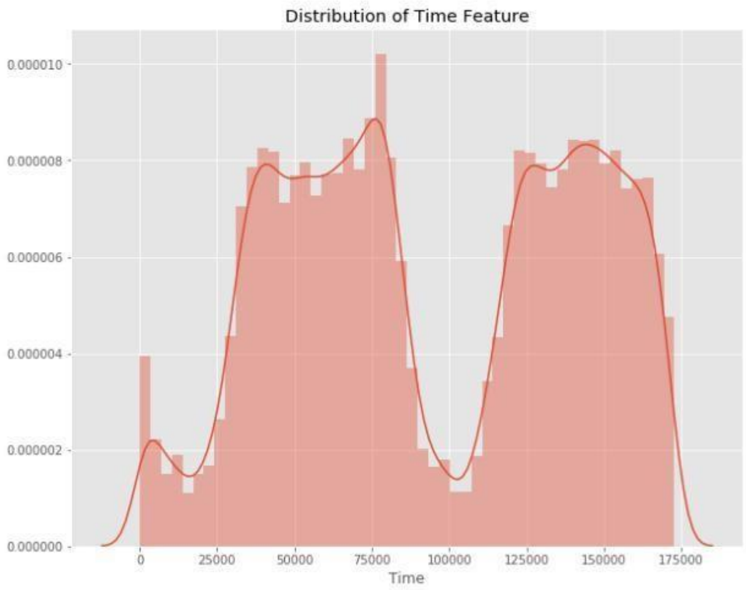

Fig. 4. Performance of the proposed method.

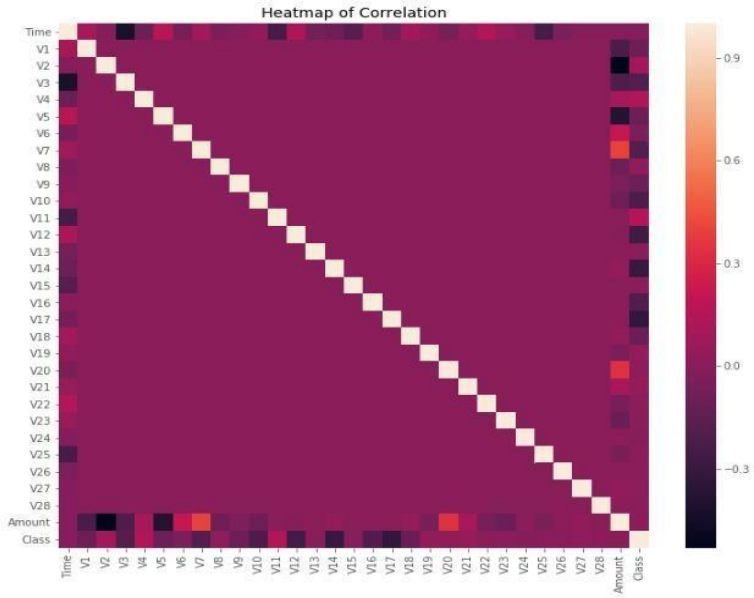

Fig. 5. Heatmap correlation.

This concept is difficult to put into practice in practice since it necessitates the collaboration of banks, which are unwilling to exchange information owing to market competition, as well as for legal concerns and the protection of their users' data. As a consequence, we searched up some reference publications that used comparable methods and gathered data. As stated in one of these reference papers: Credit card fraud is unquestionably a kind of criminal deception. This article evaluated current results in this subject and outlined the most prevalent types of fraud, as well as how to identify them. This study also goes into great depth on how machine learning may be used to improve fraud detection outcomes. Pseudo code, explanation its implementation and experimentation results.

4 | Conclusion

While the method achieves a precision of over 99.6%, when only a tenth of the data set is considered, it only achieves a precision of 28%. When the complete dataset is given into the system, however, the accuracy increases to 33%. Due to the large disparity between the number of legitimate and authentic transactions, this high percentage of accuracy is to be expected. Because the complete dataset is made up of only two days' worth of transaction records, it's only a small portion of the data that could be made

public if this research were to be utilized commercially. Because the software is built on machine learning methods, it will only get more efficient over time as more data is sent into it.

References

[1] Mohapatra, H., & Rath, A. K. (2020). *Fundamentals of software engineering: designed to provide an insight into the software engineering concepts*. BPB Publications.

[2] Mohapatra, H., Debnath, S., & Rath, A. K. (2019). Energy management in wireless sensor network through EB-LEACH. *International journal of research and analytical reviews (IJRAR)*, 56-61. https://easychair.org/publications/preprint_download/tf5s

[3] Mohapatra, H., Debnath, S., Rath, A. K., Landge, P. B., Gayen, S., & Kumar, R. (2020). An efficient energy saving scheme through sorting technique for wireless sensor network. *International journal*, *8*(8), 4278-4286.

[4] Mohapatra, H., Rath, A. K., Landge, P. B., Bhise, D. H. I. R. A. J., Panda, S., & Gayen, S. A. (2020). A comparative analysis of clustering protocols of wireless sensor network. *International journal of mechanical and production engineering research and development (IJMPERD) ISSN (P)*, *10*(3), 2249-6890.

[5] Mohapatra, H., & Rath, A. K. (2019). Fault tolerance through energy balanced cluster formation (EBCF) in WSN. *Smart innovations in communication and computational sciences* (pp. 313-321). Springer, Singapore.

[6] Mohapatra, H., & Rath, A. K. (2019). Fault tolerance in WSN through PE-LEACH protocol. *IET wireless sensor systems*, *9*(6), 358-365.

[7] Mohapatra, H., & Rath, A. K. (2021). Fault tolerance in WSN through uniform load distribution function. *International journal of sensors wireless communications and control*, *11*(4), 385-394.

[8] Mohapatra, H., & Rath, A. K. (2019). Fault-tolerant mechanism for wireless sensor network. *IET wireless sensor systems*, *10*(1), 23-30.

[9] Mohapatra, H., & Rath, A. K. (2020). Survey on fault tolerance-based clustering evolution in WSN. *IET networks*, *9*(4), 145-155.

[10] Mohapatra, H., Rath, A. K., Lenka, R. K., Nayak, R. K., & Tripathy, R. (2021). Topological localization approach for efficient energy management of WSN. *Evolutionary intelligence*, 1-11. https://link.springer.com/article/10.1007/s12065-021-00611-z

[11] Mohapatra, H., Rath, S., Panda, S., & Kumar, R. (2020). Handling of man-in-the-middle attack in wsn through intrusion detection system. *International journal*, *8*(5), 1503-1510.

[12] Mohapatra, H. (2021). *Designing of fault tolerant models for wireless sensor network* (Ph.D Dissertation, Veer Surendra Sai University of Technology). Retrieved from http://hdl.handle.net/10603/333160

[13] Nirgude, V., Mahapatra, H., & Shivarkar, S. (2017). Face recognition system using principal component analysis & linear discriminant analysis method simultaneously with 3d morphable model and neural network BPNN method. *Global journal of advanced engineering technologies and sciences*, *4*(1), 1-6.

[14] Mohapatra, H. (2009). *HCR by using neural network* (Master's Thesis, M. Tech_s Desertion, Govt. College of Engineering and Technology, Bhubaneswar).

[15] Mohapatra, H., & Rath, A. K. (2021). A fault tolerant routing scheme for advanced metering infrastructure: an approach towards smart grid. *Cluster computing*, *24*, 2193-2211.

[16] Panda, M., Pradhan, P., Mohapatra, H., & Barpanda, N. K. (2019). Fault tolerant routing in heterogeneous environment. *International journal of scientific & technology research*, *8*(8), 1009-1013.

[17] Ande, V. K., & Mohapatra, H. (2015). SSO mechanism in distributed environment. *International journal of innovations & advancement in computer science*, *4*(6), 133-136.

[18] Kumar, R., Jha, S., & Singh, R. (2020). A different approach for solving the shortest path problem under mixed fuzzy environment. *International journal of fuzzy system applications (IJFSA)*, *9*(2), 132-161.

[19] Broumi, S., Dey, A., Talea, M., Bakali, A., Smarandache, F., Nagarajan, D., ... & Kumar, R. (2019). Shortest path problem using Bellman algorithm under neutrosophic environment. *Complex & intelligent systems*, *5*(4), 409-416.

[20] Mohapatra, H. (2018). *C Programming: Practice cpp*. Kindle Edition.

[21] Kumar, R., Dey, A., Broumi, S., & Smarandache, F. (2020). A study of neutrosophic shortest path problem. *Neutrosophic graph theory and algorithms* (pp. 148-179). IGI Global.

[22] Kumar, R., Edalatpanah, S. A., Jha, S., & Singh, R. (2019). *A novel approach to solve gaussian valued neutrosophic shortest path problems*. Infinite study.

Panda and Yadlapalli| Big. Data. Comp. Vis. 1(3) (2021) 140-145

[23] Kumar, R., Edaltpanah, S. A., Jha, S., Broumi, S., & Dey, A. (2018). *Neutrosophic shortest path problem*. Infinite Study.

[24] Kumar, R., Edalatpanah, S. A., & Mohapatra, H. (2020). Note on "Optimal path selection approach for fuzzy reliable shortest path problem". *Journal of intelligent & fuzzy systems*, (Preprint), 39(5), 7653-7656.

[25] Kumar, R., Edalatpanah, S., Jha, S., Broumi, S., Singh, R., & Dey, A. A. (2019). Multi objective programming approach to solve integer valued neutrosophic shortest path problems. *Neutrosophic sets and systems*, *24*, 134-154.

[26] Kumar, R., Edalatpanah, S. A., Jha, S., & Singh, R. (2019). A Pythagorean fuzzy approach to the transportation problem. *Complex & intelligent systems*, *5*(2), 255-263.

[27] Pratihar, J., Kumar, R., Dey, A., & Broumi, S. (2020). Transportation problem in neutrosophic environment. *Neutrosophic graph theory and algorithms* (pp. 180-212). IGI Global.

[28] Pratihar, J., Kumar, R., Edalatpanah, S. A., & Dey, A. (2021). Modified Vogel's approximation method for transportation problem under uncertain environment. *Complex & intelligent systems*, *7*(1), 29-40.

[29] Gayen, S., Jha, S., Singh, M., & Kumar, R. (2019). On a generalized notion of anti-fuzzy subgroup and some characterizations. *International journal of engineering and advanced technology*, *8*(3), 385-390.

[30] Gayen, S., Smarandache, F., Jha, S., & Kumar, R. (2020). Interval-valued neutrosophic subgroup based on interval-valued triple t-norm. *Neutrosophic sets in decision analysis and operations research* (pp. 215-243). IGI Global.

[31] Gayen, S., Smarandache, F., Jha, S., & Kumar, R. (2020). *Introduction to interval-valued neutrosophic subring* (Vol. 36). Infinite Study.

[32] Gayen, S., Smarandache, F., Jha, S., Singh, M. K., Broumi, S., & Kumar, R. (2020). Introduction to plithogenic hypersoft subgroup. *Neutrosophic sets and system, 33*, 14-22.

[33] Gayen, S., Smarandache, F., Jha, S., Singh, M. K., Broumi, S., & Kumar, R. (2020). Introduction to plithogenic subgroup. *Neutrosophic graph theory and algorithms* (pp. 213-259). IGI Global.

[34] Gayen, S., Smarandache, F., Jha, S., Singh, M. K., Broumi, S., & Kumar, R. (2020). *Soft subring theory under interval-valued neutrosophic environment* (Vol. 36). Infinite Study.

[35] Kumar, R., Edalatpanah, S. A., Jha, S., Gayen, S., & Singh, R. (2019). Shortest path problems using fuzzy weighted arc length. *International journal of innovative technology and exploring engineering*, *8*(6), 724-731.

[36] Kumar, R., Jha, S., & Singh, R. (2017). Shortest path problem in network with type-2 triangular fuzzy arc length. *Journal of applied research on industrial engineering*, *4*(1), 1-7.

[37] Mohapatra, H., Panda, S., Rath, A., Edalatpanah, S., & Kumar, R. (2020). A tutorial on powershell pipeline and its loopholes. *International journal of emerging trends in engineering research*, *8*(4), 975-982.

[38] Kumar, R., Edalatpanah, S. A., Gayen, S., & Broum, S. (2021). Answer Note "A novel method for solving the fully neutrosophic linear programming problems: Suggested modifications". *Neutrosophic sets and systems*, *39*(1), 148-152. https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1751&context=nss_journal

[39] Panda, H., Mohapatra, H., & Rath, A. K. (2020). WSN-based water channelization: an approach of smart water. *Smart cities—opportunities and challenges*, *58*, 157-166.

[40] Mohapatra, H., & Rath, A. K. (2020, October). Nub less sensor based smart water tap for preventing water loss at public stand posts. *2020 IEEE microwave theory and techniques in wireless communications (MTTW)* (Vol. 1, pp. 145-150). IEEE.

[41] Rout, S. S., Mohapatra, H., Nayak, R. K., Tripathy, R., Bhise, D., Patil, S. P., & Rath, A. K. (2020). Smart water solution for monitoring of water usage based on weather condition. *International journal*, *8*(9), 5335-5343.

[42] Mohapatra, H., & Rath, A. K. (2021). IoE based framework for smart agriculture. *Journal of ambient intelligence and humanized computing*, 1-18. https://doi.org/10.1007/s12652-021-02908-4

[43] Mohapatra, H., & Rath, A. K. (2021). An IoT based efficient multi-objective real-time smart parking system. *International journal of sensor networks*, *37*(4), 219-232.

[44] Mohapatra, H., & Rath, A. K. (2019). Detection and avoidance of water loss through municipality taps in India by using smart taps and ICT. *IET wireless sensor systems*, *9*(6), 447-457.

[45] Mohapatra, H. (2020). Offline drone instrumentalized ambulance for emergency situations. *IAES International journal of robotics and automation*, *9*(4), 251-255.